

Complete genome sequence of *Limnothrix* sp. CMB-01, a self-flocculating cyanobacterium isolated from rare earth element wastewater

Shanren Li,¹ Dandan Che,¹ Ming Li,² Zhefu Li,¹ Xibin Qiu,¹ Mingzi Wang¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genome sequence of *Limnothrix* sp. CMB-01, a self-flocculating cyanobacterium isolated from rare earth element wastewater. The assembled genome comprises a single circular chromosome of 4,574,928 bp with a GC content of 55.22%, providing genomic insights into the metabolic potential and adaptive traits of this underexplored cyanobacterium.

KEYWORDS *Limnothrix*, complete genome, self-flocculating cyanobacterium

Wastewater generated from rare earth element (REE) production exhibits high ammonia nitrogen concentrations, underscoring the urgent need for efficient and sustainable treatment strategies. Self-flocculating microalgae offer a promising bioremediation approach, enabling effective ammonia nitrogen removal and low-energy harvesting through natural aggregation and sedimentation (1–3). Although cyanobacteria of the genus *Limnothrix* are widely distributed in aquatic environments, genomic resources for strains with bioremediation potential remain limited. Strain CMB-01 was isolated from REE production wastewater collected in Changting County, Longyan City, Fujian Province, China (25.74°N, 116.36°E). Isolation was performed by serially diluting the wastewater sample with sterile water, plating onto BG11 agar plates, and incubating at 25 °C for 10 days under continuous illumination (75 $\mu\text{mol m}^{-2} \text{s}^{-1}$). A single colony was subsequently purified on BG11 plates.

Genomic DNA of strain CMB-01 was sequenced using a hybrid approach combining Oxford Nanopore (long-read) and DNBSEQ (short-read) platforms at Sangon Biotech (Shanghai, China). The strain was cultured in BG11 medium under continuous illumination (75 $\mu\text{mol m}^{-2} \text{s}^{-1}$) with shaking at 110 rpm and 25°C. DNA was extracted using the NucleoBond HMW DNA kit (Macherey-Nagel, Germany) and quantified using a NanoDrop One spectrophotometer and Qubit 4.0 fluorometer (Thermo Fisher Scientific, USA). DNA integrity and purity were assessed by 0.75% agarose gel electrophoresis.

For long-read sequencing, a DNA library was prepared with the SQK-LSK110 ligation sequencing kit without DNA shearing and sequenced on a PromethION R9.4.1 flow cell. Basecalling, barcode segmentation, and adapter trimming were performed using Guppy v6.5.6 (4) in high-accuracy mode, generating 95,402 raw reads (average length = 10,551 bp; N_{50} = 21,764 bp). Raw reads were filtered using Filtlong v0.2.1 (5) to remove short (<1,000 bp) and low-quality reads ($Q < 6$), resulting in 88,009 subreads (average length = 11,294 bp; N_{50} = 21,831 bp). For short-read sequencing, DNA was sheared to approximately 300 bp, and a 150 bp paired-end library was prepared using the MGI Easy DNA Library Prep Kit and sequenced on a DNBSEQ-T7 instrument, generating 39,408,506 raw reads (5.91 Gb total). Raw reads were processed using fastp v0.23.0 (6) for adapter trimming and quality filtering, resulting in 39,404,298 clean reads (5.88 Gb total).

Editor John J. Dennehy, Queens College
Department of Biology, Queens, New York, USA

Address correspondence to Mingzi Wang,
mingziw@fjnu.edu.cn.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 16 February 2026

Accepted 11 March 2026

Published 23 March 2026

Copyright © 2026 Li et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Genome assembly and annotation statistics for *Limnothrix* sp. CMB-01

Parameter	Value
GenBank accession number	CM143720
Genome size (bp)	4,574,928
GC content (%)	55.22
N_{50} (bp)	4,574,928
Coverage (×)	894.73
Genes (total)	3,789
CDSs (total)	3,735
CDSs (with protein)	3,668
CDSs (without protein)	67
rRNA genes (5S, 16S, 23S)	6 (2, 2, 2)
tRNA genes	44
Noncoding RNAs	4
Pseudogenes	68

Filtered long reads were assembled using Canu v2.2 (7), and the resulting contig was circularized with Circlator v1.5.5 (8). Gap filling and sequence correction were performed by integrating short reads using GapFiller v1.11 (9) and PrInSeS-G v1.0.0 (10), respectively. The assembly was iteratively polished with Pilon v1.22 (11) until no further errors were detected. CheckM v1.2.4 (12) estimated genome completeness at 98.28% with 0.16% contamination, and average genome coverage was approximately 894.73×.

The final assembly yielded a single circular chromosome of 4,574,928 bp with a G + C content of 55.22% (unrotated). Gene annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline v6.10 (13). The 16S rRNA gene sequence showed 99.00% identity to known *Limnothrix* species, suggesting that strain CMB-01 represents a novel strain within the genus. Default parameters were used for all software unless otherwise noted. The principal genomic features are summarized in Table 1.

ACKNOWLEDGMENTS

This work was supported by the Fujian Provincial Science and Technology Major Project (2023YZ037002) and the National Natural Science Foundation of China (no. 32471541).

AUTHOR AFFILIATIONS

¹National and Local United Engineering Research Center of Industrial Microbiology and Fermentation Technology, College of Life Sciences, Fujian Normal University, Fuzhou, China

²Appraisal Center for Environment and Engineering, Ministry of Ecology and Environment of China, Beijing, China

AUTHOR ORCID*s*

Shanren Li  <http://orcid.org/0000-0001-9302-7998>

Mingzi Wang  <http://orcid.org/0000-0001-6230-6601>

FUNDING

Funder	Grant(s)	Author(s)
Fujian Provincial Department of Science and Technology	2023YZ037002	Mingzi Wang
National Natural Science Foundation of China	32471541	Mingzi Wang

AUTHOR CONTRIBUTIONS

Shanren Li, Data curation, Formal analysis, Investigation, Writing – original draft | Dandan Che, Data curation, Formal analysis, Methodology | Ming Li, Formal analysis, Investigation, Resources | Zhefu Li, Formal analysis, Investigation, Resources | Xibin Qiu, Formal analysis, Investigation, Resources | Mingzi Wang, Conceptualization, Funding acquisition, Project administration, Writing – review and editing

DATA AVAILABILITY

The complete genome sequence of *Limnothrix* sp. CMB-01 has been deposited in NCBI GenBank under accession number [CM143720](#) (BioProject number [PRJNA1420451](#) and BioSample number [SAMN55230565](#)). The raw sequence data are available in the Sequence Read Archive under accession numbers [SRR37182836](#) (ONT) and [SRR37182837](#) (DNBSEQ), respectively.

REFERENCES

- Che D, Lai Y, Weng Z, Li M, Huang G, Zheng M, Wang M. 2024. Self-flocculating *Spirulina platensis* CMB-02 to efficiently treat ammonia nitrogen of rare earth elements wastewater. *Bioresour Technol* 411:131360. <https://doi.org/10.1016/j.biortech.2024.131360>
- Zhou Y, He Y, Guo X, Dai J, Lai X, Hong B, Chen B, Wang M. 2023. Pilot-scale remediation of rare earth elements ammonium wastewater by *Chlamydomonas* sp. YC in summer under outdoor conditions. *Bioresour Technol* 372:128674. <https://doi.org/10.1016/j.biortech.2023.128674>
- Zhou Y, He Y, Zhou Z, Xiao X, Wang M, Chen B. 2022. A newly isolated microalga *Chlamydomonas* sp. YC to efficiently remove ammonium nitrogen of rare earth elements wastewater. *J Environ Manage* 316:115284. <https://doi.org/10.1016/j.jenvman.2022.115284>
- Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>
- Wick R. 2017. Filtlong. Available from: <https://github.com/rrwick/Filtlong>
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>
- Massouras A, Hens K, Gubelmann C, Uplekar S, Decouttere F, Rougemont J, Cole ST, Deplancke B. 2010. Primer-initiated sequence synthesis to detect and assemble structural variants. *Nat Methods* 7:485–486. <https://doi.org/10.1038/nmeth.f.308>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>